

# Big Data

## Big Data Analytics: Driving Value Beyond the Hype

OCTOBER 2, 2012

CAMBRIDGE, MASSACHUSETTS

WE ARE IN THE MIDST OF A BIG DATA EXPLOSION, in volume and velocity. The challenges for Federal agencies, industry, and other organizations are to figure out how to manage and analyze Big Data to create value, increase accurate predictions, and provide approaches that drive sound decision making.

The U.S. Department of Transportation's Volpe Center and the U.S. Air Force Life Cycle Management Center, Hanscom Air Force Base, co-hosted a roundtable of experts to discuss the topic and found that with these great challenges, new insights and opportunities emerge—new technologies and processes are being considered, industry trends and best practices are being reviewed, and deep data is waiting to be mined. It is projected that analyzing Big Data—will become a key basis of competition, enabling productivity, growth, and innovation as long as the right policies and enablers are in place.

Significant workforce challenges lie ahead as well. A workforce with deep analytical experience will be needed and managers will need to become data literate. Those who figure out how to unite domain expertise with data science will move to the forefront.

“Big Data” refers to enormous volumes of information.

Big Data is being generated by sources within the IT infrastructure, smartphones, personal computers, digital photography, streaming media, GPS, real-time control systems in automobiles, robot vehicles ... and the list is growing.

[www.volpe.dot.gov](http://www.volpe.dot.gov)



U.S. Department of Transportation

**Research and Innovative Technology Administration**

John A. Volpe National Transportation Systems Center

## The Program

*Setting the Stage*

*Working Smarter:  
Implications of Ubiquitous Data*

*Predictive Analytics: Delivering Insight  
to Decision Makers*

*Improving Transportation Safety and  
Operations through a Big Data Ecosystem*

*Moderated Discussion*

## The Expert Panel

**Robert C. Johns**, Associate Administrator and  
Director of the Volpe Center

**Dr. Anne Quadgras**, Research Scientist,  
Massachusetts Institute of Technology  
Sloan School, Center for Information  
Systems Research

**Frank Stein**, Director, Analytics Solutions Center,  
Office of the Chief Technology Officer, IBM Federal

**Peter Bostrom**, Transformation Director,  
Department of Defence (DoD), Engineered Systems,  
Oracle Public Sector

**Tim Schmidt**, Acting Chief Information Officer,  
U.S. Department of Transportation

*Please note: The views of non-DOT participants do not necessarily reflect the views of the  
U.S. Department of Transportation.*



## Big Data Is Big News

*“The focus is on Big Data. The title: ‘Big Data Analytics:  
Driving Value Beyond the Hype’ is appropriate. We should take advantage of  
new opportunities: Big Data is big news.”*

– **Robert C. Johns, Associate Administrator and Director**  
Volpe National Transportation Systems Center

- Beyond the catch phrase, Big Data is defined by the enormous volume of data, its diversity and variety, and the speed by which it can be acquired and queried.
- We are experiencing a data explosion. With corporate data nearly doubling each year, companies process about 1,000 times more data than a decade ago. The Federal government has more big data than it knows what to do with, and it’s only going to get bigger.
- Big Data creates value and unlimited opportunities in many ways: making information usable at a faster frequency, substantially improving decision-making, increasing accurate predictions, and allowing precisely tailored products and services for specific customer segments.
- The White House Big Data R&D initiative is creating a vision at Federal and private levels. An issue for all of us calls for the challenge to expand the workforce needed to develop and use big data technologies. Goals also include advancement of technologies, and to accelerate the pace of capturing value from the data.



## How We Drive Value from Data

*“What’s different now that data is ubiquitous? What we’re really talking about is ‘Driving Value Beyond the Hype’. There’s not yet a lot of boundaries — slow down and see how we drive value from data.”*

— **Dr. Anne Quaadgras**, Research Scientist, Massachusetts Institute of Technology Sloan School, Center for Information Systems Research

- Using ubiquitous data well is critical. It means working smarter by using “little” data that has become abundant to enhance both operational and strategic decision making.
- Getting value from data is hard—it involves changing culture. “An Evidence Based Management Culture” is what really makes a difference that allows for a platform where there is common data and common technologies and systems communicate with each other. The goal of the smart organization is to use data from the platform to optimize roles and what people are doing.
- Top performers are working smarter, and in companies surveyed that ranked 70% at the top quartile, they answered true to characteristics of working smarter: Empower operational decision makers with useful information; create and revise business rules based on business analytics; automate repetitive business processes; provide external partners with timely access to information; and rely on a single source of truth for data.
- Foundational practices to working smarter consist of four important concepts: **Scorecards**—Ask, what am I doing today? How can I get better?; **Business Rules**—These are specified actions to meet strategic objectives; **Pivotal Managers**—People who manage those who generate revenue; and **Coaching**—Have regular conversations with front line people (metrics) with focused feedback that staff can use to measure and improve successes.
- Management imperatives for working smarter are part of the information backbone—to develop a backbone, manage business rules and decisions, and redesign structures, roles, and accountability.
- Working smarter is an ongoing process that requires a focus on daily operations: define and integrate important data, test and optimize business rules, coach pivotal managers, refine scorecards, and cultivate talents in people.

---

## Watson and Big Data

*“Can we design a computing system that rivals a human’s ability to answer questions posed in natural language, interpreting meaning and context and retrieving, analyzing and understanding vast amounts of information in real-time?”*

– **Frank Stein**, Director, Analytics Solutions Center, Office of the CTO, IBM Federal



- Watson’s real value proposition: efficient decision support over unstructured and structured content. While primarily an IBM effort, Watson’s development team includes faculty and students from Carnegie Mellon University, University of Massachusetts Amherst, University of Southern California’s Information Sciences Institute, University of Texas at Austin, Massachusetts Institute of Technology, University of Trento, and Rensselaer Polytechnic Institute. *(Continued on next page.)*

- Huge volumes of data are coming at us from many sources in real time—it's more instantaneous than before. Big data is now mostly 85% unstructured, and only 15% is structured, making it a challenge to sift through and properly analyze to get value and usage out of that data. Data in motion is 10,000 times faster; real time is where we're going.
- In addition to volume, velocity, and variety, there is a new dimension of Big Data: 'veracity' (or truth). There is doubt or uncertainty in data in lots of ways: processes that contain 'randomness' (uncertain travel times), data input with text entry typos or GPS fuzziness, or poor forecasting with all of predictive modeling only approximate. Uncertainty from many sources is becoming a bigger issue, and there is a need to start building systems that carry "confidence intervals" along through the calculations.
- Organizations do not do a good job at analytics today, and business leaders are finding they don't have the data they need to make good decisions. New Big Data brings opportunities and requires new analytics. There must be a change in the paradigm for Big Data to work—by moving analytics to the data; this changes technology. More companies are starting to use analytics data to improve.
- How do you process data in real time? The Big Data platform designed for variety and to analyze many data types simultaneously is "Stream Computing." The stream computing process takes data and processes it "on the fly" before the data hits the disk.
- Is Watson a Big Data Analytical System? Yes. The Watson system is an approach to use against Big Data ... it represents a major leap forward in natural language processing, analytics, and system design."

---

## It's All About the Numbers

*"What's next in transportation safety? Approximately 75% of aviation accidents are due to human error. Without integration, do we improve persons in the loop, or machine-to-machine end ... quantify, and then at the end of the day, it's all about the numbers."*



– **Peter Bostrom**, Transformation Director, DoD, Engineered Systems, Oracle Public Sector

- Working with Big Data requires a new specialized infrastructure and skill sets. The challenge is a time and cost-to-value ratio. Even with the right tools, how fast can we get there, at what cost, and will it make a difference?
- Big Data can be structured, semi-structured, and unstructured, which poses the problem of how to integrate these enormous amounts of small pieces of information. Big Data requires new solutions, and some organizations are using techniques borrowed from data mining; it's dynamically decided. The future lies in integration of large scaled pieces of information.
- There are a number of projects and programs underway where agencies and organizations are building Big Data ecosystems. For example, persistent Intelligence, Surveillance and Reconnaissance is moving analytics to where the sensors are and determining what is important in order to solve too much data coming across a too narrow pathway for reaching the decision maker.
- Very little data is totally unstructured. That's why the internet is so important to the Washington D.C. area; there's an understanding of infrastructure and connectivity. D.C. has the highest numbers in education levels and expertise. These are strengths—each organization has to think of ways to identify strengths, bring the most value—using the least amount of money in the least amount of time.



## Moderated Discussion

During the question and answer period, U.S. DOT's Acting Chief Information Officer posed:

*“What are the Federal challenges we are facing today? Not so much that there is one silver bullet of best practices, but rather ... can we use data known to us to pinpoint who needs what? How do we find data sets and leverage them? How do we get Big Data down to manageable data? What are some common challenges? What are some examples of best practices?”*

– **Tim Schmidt**, Acting Chief Information Officer, U.S. Department of Transportation

- **Dr. Anne Quaadgras**, MIT Sloan School, “How do we do business ethically with the data we have? The government’s function is to decide what data is really needed. Ask, how can we do this? How much money is needed to put into this? Data storage takes care of Health Insurance Portability and Accountability Act (HIPPA) privacy issues, and it should be a transparent government process to decide how much data to use and how it should be used. Ask the question, what are we using this data for? Real-time changes the rules and requires an ongoing check of metrics. It used to be transformation would end at some point—now it’s continuous.  
Also with companies putting in enterprise resource planning (ERP), how do we use the sea of data coming out of ERPs? This is the next step—get at the data and not enter another ERP into the pipeline. This is a key point: “make investments in the people, not the pipe.”
- **Dr. Tim Rudolph**, Hanscom Air Force Base, “With the Department of Defense there is a gradual change from net-centric to data centric; the value of real time. Data analytics is getting very big. Some data are saved longer for forensic purposes. ERP planning is costly.”
- **Frank Stein**, IBM Federal, “With IBM, ERPs lock the data; we’re trying to unlock the data. We need to move away from system-centricity. Where are priorities in real time? There are problems with low probability. How do we justify budgets? How do we optimize—constantly build a learning system.”
- **Dr. Mark Linderman**, Air Force Research Laboratories, “Finance and contracting have legacy systems with analytics on top of them. There is no automated way to get at the data. We need fundamental ways for the data to support all systems.”
- **Tim Schmidt**, U.S. DOT, “Data sets are everywhere with no way to get at it. Safety data alone—there are thousands of data sets. We need to build applications to leverage these data sets and lower costs.”
- **Ann Aylward**, Volpe/U.S. DOT, “Figure out how to use international freight and freight movement. Systemwide data is available and dates back. There is a real disconnect in the existence of data and the ability to use it. An example would be 2003 data in a 2012 report—this is the most recent nationwide data available. We need to get fresh industry data.”
- **Peter Bostrom**, Oracle Public Sector, “In the medical community there are challenges—to figure out ways sensor data is being collected: from data sensor and video feeds to how to multi-task. There’s a continuous process for improvement. Figure out how to disseminate talents of people; get a bigger bang for the buck, and a greener America by looking at collaboration and improving expertise.”

*(Continued on next page.)*

- **Stephen Popkin**, Volpe/U.S. DOT, “When talking about public tolerance, data is not going to be perfect. Data can say anything; Big Data can really say anything. Is there tolerance for ambiguity? How is it viewed?”
- **Frank Stein**, IBM Federal, “The National Weather Service is improving perception with probabilities. It gives a range of possibilities—this carries over to other aspects like transportation; for example, estimated arrival times. There’s a need to better understand data analytics.”

---

## Meet Volpe

Volpe has been helping the transportation community navigate the most challenging problems for more than 40 years. As the National Transportation Systems Center, our mission is to improve transportation by anticipating and addressing emerging issues and advancing technical, operational, and institutional innovations across all modes. Part of the U.S. Department of Transportation’s Research and Innovative Technology Administration, Volpe is a unique Federal agency that is 100 percent funded by sponsor projects.

Home to renowned multidisciplinary expertise in all modes of transportation, Volpe serves its sponsor agencies with advanced technologies, research, and programs to ensure a fast, safe, efficient, accessible, and convenient transportation system that meets vital national and international interests and enhances the quality of life for the traveling public, today and into the future.

---

## About the Colloquia Series

Volpe, The National Transportation Systems Center, is pleased to present a Colloquia Series on Transportation Challenges and Opportunities. The series, which brings together industry experts from government, academia, and the private sectors, continues Volpe’s long tradition of facilitating knowledge exchange across the transportation community and takes a fresh approach in addressing today’s transportation challenges and issues. The series is available via webinar and members of the transportation community are encouraged to participate in question and answer periods.

*Please join us:*

<http://www.volpe.dot.gov/coi/outreach/colloquia/index.html>

<http://www.volpe.dot.gov/coi/outreach/source/index.html>

<http://www.volpe.dot.gov/coi/outreach/trajectories/index.html>

For more information, please contact: Ellen E. Bell, Director of Strategic Initiatives for Research and Innovation at the Volpe Center: [Ellen.Bell@dot.gov](mailto:Ellen.Bell@dot.gov)



U.S. Department of Transportation

**Research and Innovative Technology Administration**

John A. Volpe National Transportation Systems Center